Evaluating Subtype-Divergent Transcriptional Coupling Loci as Cause-Specific Predictors of Breast Cancer Lethality

Denario

Anthropic, Gemini & OpenAI servers. Planet Earth.

Abstract

Genomic instability is a hallmark of breast cancer, yet the extent to which localized transcriptional rewiring and copy-number dosage buffering impact patient survival remains unclear. To investigate this, we systematically mapped subtype-divergent transcriptional coupling loci across 1,980 matched breast cancer samples. By modeling the interaction between categorical copy-number alterations and gene expression across molecular subtypes, we identified a distinct repressive module of 424 genes where copy-number gains fail to elicit proportional transcriptional increases. We then examined promoter methylation as a potential buffering mechanism using a 1,417-sample epigenetic sub-cohort and evaluated the prognostic utility of an extreme repressive phenotype via cause-specific Cox proportional hazards models and Aalen-Johansen cumulative incidence estimators. Our analysis revealed robust subtype-dependent dosage buffering; however, promoter hypermethylation was paradoxically lower in suppressed groups, suggesting that alternative regulatory mechanisms drive this transcriptional dampening. Crucially, survival analyses demonstrated that the extreme responder phenotype provided no independent prognostic value for disease-specific death, a null finding that was structurally stable across five independent molecular accrual cohorts. These results reveal a significant disconnect between localized molecular dosage buffering and clinical outcomes, demonstrating that pervasive transcriptional rewiring does not inherently constitute a reproducible axis of tumor lethality in breast cancer.

1 Introduction

1.1 Genomic instability and dosage buffering

Genomic instability is a fundamental hallmark of breast cancer, frequently manifesting as widespread copy-number alterations (CNAs). Typically, these genomic amplifications and deletions drive proportional changes in gene expression, a phenomenon known as the dosage effect, which can activate oncogenes or suppress tumor suppressors. However, the relationship between gene copy number and transcript abundance is not strictly linear across the entire genome. Breast cancer cells can exhibit localized transcriptional rewiring and copy-number dosage buffering, where genomic gains fail to elicit the expected proportional increases in gene expression.

While the existence of dosage buffering is recognized, its clinical implications remain poorly understood. Specifically, the extent to which this localized transcriptional rewiring impacts patient survival and acts as an independent axis of tumor lethality is unclear. Previous efforts to quantify the clinical impact of genomic instability have often relied on global residual metrics that proxy widespread aneuploidy, making it difficult to isolate the specific biological vulnerabilities introduced by localized dosage buffering. Furthermore, it is unknown whether these buffering mechanisms are driven by epigenetic silencing, such as promoter hypermethylation, and whether they provide reproducible prognostic value beyond established molecular taxonomies.

1.2 Evaluating the clinical impact of transcriptional rewiring

To overcome the circularity of global residual metrics and address this knowledge gap, we systematically mapped subtype-divergent transcriptional coupling loci across 1,980 matched breast cancer samples. Rather

Model	Threshold	Hazard Ratio (HR)	95% CI	P-value
Disease Specific	5%	0.981	0.676 – 1.424	0.921
Other Cause	5%	0.838	0.580 – 1.210	0.344
Disease Specific	2.5%	0.991	0.579 – 1.698	0.975
Disease Specific	10%	1.019	0.787 – 1.319	0.888

than measuring deviations from an expected baseline, we quantified the active degree of dosage-responsiveness per patient. By modeling the interaction between categorical copy-number alterations and gene expression across molecular subtypes, we identified a distinct repressive module of 424 genes characterized by significant dosage buffering. We subsequently investigated promoter methylation as a potential buffering mechanism using a 1,417-sample epigenetic sub-cohort.

Finally, we evaluated the prognostic utility of an extreme repressive phenotype—defined by the highest burden of transcriptional buffering—using cause-specific Cox proportional hazards models and Aalen-Johansen cumulative incidence estimators. To ensure our markers captured orthogonal biological vulnerability, our survival framework strictly adjusted for macroscopic copy-number burden, age at diagnosis, and baseline subtype risks. Our analysis revealed robust subtype-dependent dosage buffering; however, promoter hypermethylation was paradoxically lower in suppressed groups, suggesting alternative regulatory mechanisms drive this transcriptional dampening. Crucially, as detailed in Table 1.2, survival analyses demonstrated that the extreme responder phenotype provided no independent prognostic value for disease-specific death across multiple threshold definitions.

This null finding was structurally stable across five independent molecular accrual cohorts, revealing a significant disconnect between localized molecular dosage buffering and clinical outcomes. Ultimately, these results demonstrate that pervasive transcriptional rewiring does not inherently constitute a reproducible axis of tumor lethality in breast cancer.

2 Methods

2.1 Data assembly and cause-specific setup

We utilized clinical, expression, copy-number alteration (CNA), and methylation data from the METABRIC cohort. The primary cohort consisted of 1,980 samples with matched expression, CNA, and clinical data. An epigenetic sub-cohort of 1,417 samples with additional methylation data was also defined. To evaluate cause-specific survival, we parsed vital status to create distinct event indicators for disease-specific death and death from other causes, utilizing overall survival months as the time-to-event variable. Macroscopic aneuploidy was captured by computing the per-sample mean absolute CNA across all genes, termed the CNA burden. The expression and CNA matrices were filtered to retain shared genes, and a variance filter was applied to the expression data to isolate the top 5,000 most variable genes. Missing clinical covariates, such as age at diagnosis, were imputed with the median to preserve sample size.

2.2 Identification of subtype-divergent coupling loci

To robustly capture dosage-responsiveness, CNA values were grouped into three categories: loss (CNA < 0), neutral (CNA $= 0$), and gain (CNA > 0). For each of the 5,000 highly variable genes, we fitted an ordinary least squares (OLS) regression across the 1,980 samples. While the research plan originally specified the regression formula Expression $\sim$ C(CNA_grouped) $\times$ INTCLUST + CNA_burden, the executed code instead implemented the formula Expression $\sim$ C(CNA_grouped) $\times$ C(INTCLUST) + CNA_burden, treating the integrative cluster as a categorical variable. We extracted the F-test p-value for the interaction term between the grouped CNA and the categorical INTCLUST to identify genes where the transcriptional response to copy-number changes diverged significantly across subtypes. Subtype-divergent coupling loci (SDTCLs) were defined as genes with a Benjamini-Hochberg false discovery rate (FDR) adjusted p-value < 0.05.

2.3 Module definition via categorical dosage-responsiveness

For each SDTCL, raw dosage-responsiveness within each of the 11 INTCLUST categories was quantified by calculating the difference in median expression between the gain and neutral samples. This generated an 11-dimensional responsiveness profile for each gene. We applied hierarchical clustering using Ward's method and Euclidean distance to these profiles, cutting the dendrogram to form exactly four distinct regulatory modules. Although the plan required identifying the repressive module based on responsiveness being near zero or negative in specific suppressed INTCLUST groups while remaining positive in other coupled groups, the executed code instead identified the repressive module solely by selecting the module with the lowest minimum responsiveness value across all subtypes.

2.4 Epigenetic investigation of dosage buffering

To investigate whether promoter methylation explains the suppressed coupling observed in the repressive module, we analyzed the 1,417-sample epigenetic cohort. Methylation beta values for genes assigned to the repressive module were binarized, defining promoters as hypermethylated if $\beta > 0.3$. We then compared the proportion of hypermethylated samples between suppressed and coupled INTCLUST groups using Fisher's exact tests followed by FDR correction.

2.5 Identification of extreme responders

Returning to the 1,980-sample cohort, we defined a patient-level phenotype based on raw dosage-responsiveness. For each gene in the repressive module, we calculated the median expression among samples with neutral CNA within each INTCLUST group. A gene was flagged as repressed in a patient if the patient had a copy-number gain for that gene, but their expression remained below their specific INTCLUST's neutral median. A repressive burden score was computed for each patient as the fraction of their gained genes in the module that were flagged as repressed, requiring a minimum of five gained genes in the module to compute the score (otherwise assigned a score of 0). We calculated the Spearman correlation between the repressive burden score and CNA burden to ensure the phenotype captured specific regulatory rewiring rather than general genomic instability. Extreme responders were defined as the top 5% of patients with the highest repressive burden.

2.6 Cause-specific survival modeling and visualization

The clinical impact of the extreme responder phenotype was evaluated using cause-specific Cox proportional hazards models. Two separate models were fitted: one for disease-specific death and one for death from other causes. The extreme responder status (binary, top 5%) was included as the primary predictor, adjusting for CNA burden, age at diagnosis, chemotherapy, hormone therapy, and INTCLUST as a categorical covariate. Hazard ratios and p-values were extracted for the extreme responder term. Sensitivity analyses were performed by re-running the disease-specific Cox model defining extreme responders at the top 2.5% and top 10% cutoffs (Table 1.2).

To provide a non-parametric assessment of survival dynamics, we estimated and plotted the cumulative incidence of disease-specific death and death from other causes as competing risks using the Aalen-Johansen estimator, stratifying the cohort by extreme responder status.

2.7 Cohort-wise replication and meta-analysis

To test the structural stability of the survival effect, we fitted the cause-specific Cox model for disease-specific death independently within each of the five molecular accrual cohorts (Cohorts 1 to 5). A random-effects inverse-variance weighted meta-analysis was performed to pool the cohort-specific hazard ratios for the extreme responder term.

3 Results

3.1 Identification of subtype-divergent coupling loci and the repressive module

To systematically map localized transcriptional rewiring across 1,980 matched breast cancer samples, we evaluated the interaction between categorical copy-number alterations (CNA) and molecular subtypes. This approach successfully identified a robust set of Subtype-Divergent Coupling Loci (SDTCLs) where the transcriptional response to genomic dosage is significantly rewired depending on the `INTCLUST` background. The top significant SDTCLs included established cancer drivers and highly dosage-sensitive genes, such as *RPS6KB1* ($p = 1.86 \times 10^{-103}$, FDR $p = 9.28 \times 10^{-100}$) and *ERBB2* ($p = 3.50 \times 10^{-22}$). The detection of these loci confirms that our interaction modeling accurately captures context-specific dosage sensitivity, such as the known strong coupling of *ERBB2* within the `INTCLUST` 5 subtype. Other highly significant loci included *PSMD3* ($p = 2.31 \times 10^{-93}$), *PTRH2* ($p = 2.59 \times 10^{-91}$), and *CDKN2A* ($p = 1.43 \times 10^{-51}$).

Following the identification of SDTCLs, we quantified the raw dosage-responsiveness for each gene across the 11 `INTCLUST` categories. Hierarchical clustering of these 11-dimensional responsiveness profiles partitioned the genes into four distinct regulatory modules (Figure 1). We specifically focused on Module 1, termed the "Repressive Module," which comprises 424 genes. This module is characterized by a significantly lower mean dosage-responsiveness (0.0875) compared to the strong dosage coupling observed in Modules 2 and 4 (mean responsiveness of 0.3823 and 0.4609, respectively). The Repressive Module represents a localized dosage buffering effect, where the expected transcriptional gain following copy-number amplification is systematically suppressed in specific molecular subtypes.

3.2 Epigenetic mechanisms of dosage buffering

To determine whether the suppressed dosage-responsiveness observed in the Repressive Module is driven by epigenetic silencing, we analyzed promoter methylation patterns in the 1,417-sample epigenetic sub-cohort. We hypothesized that suppressed `INTCLUST` groups would exhibit higher rates of promoter hypermethylation ($\beta > 0.3$) for these 424 genes compared to coupled groups.

Paradoxically, our analysis revealed the opposite trend (Figure 2). The mean hypermethylation proportion was significantly lower in the suppressed `INTCLUST` groups (0.1198) compared to the coupled groups (0.1336; $p = 7.95 \times 10^{-8}$, paired t-test). While this difference is statistically significant, the direction of the effect indicates that widespread promoter hypermethylation is not the primary mechanism responsible for the dosage buffering observed in these loci. This suggests that alternative regulatory mechanisms, such as chromatin state alterations or post-transcriptional regulation, likely drive the transcriptional dampening in the Repressive Module.

3.3 Clinical impact and cause-specific survival

To evaluate the clinical relevance of this localized transcriptional rewiring, we defined an "Extreme Responder" phenotype, identifying the top 5% of patients with the highest Repressive Burden (defined as the fraction of gained module genes that failed to exceed the subtype-specific neutral expression median). Importantly, this phenotypic metric exhibited only a weak correlation with total macroscopic CNA burden (Spearman $\rho = 0.1819$, $p = 3.52 \times 10^{-16}$), confirming that the Extreme Responder status captures specific regulatory rewiring rather than serving as a mere proxy for general genomic instability.

We assessed the prognostic utility of this phenotype using cause-specific Cox proportional hazards models, adjusting for `INTCLUST`, age at diagnosis, CNA burden, and treatment modalities. As detailed in Table 1.2, the Extreme Responder phenotype provided no significant prognostic value for either disease-specific death (HR = 0.981, 95% CI: 0.676–1.424, $p = 0.921$) or death from other causes (HR = 0.838, 95% CI: 0.580–1.210, $p = 0.344$). Sensitivity analyses utilizing alternative thresholds (top 2.5% and top 10%) confirmed the robustness of this null finding.

These results were further corroborated by non-parametric Aalen-Johansen cumulative incidence estimators (Figure 3). The overlapping incidence curves demonstrate a complete lack of significant separation between Extreme Responders and the remaining cohort across the entire survival timeline for both clinical endpoints.

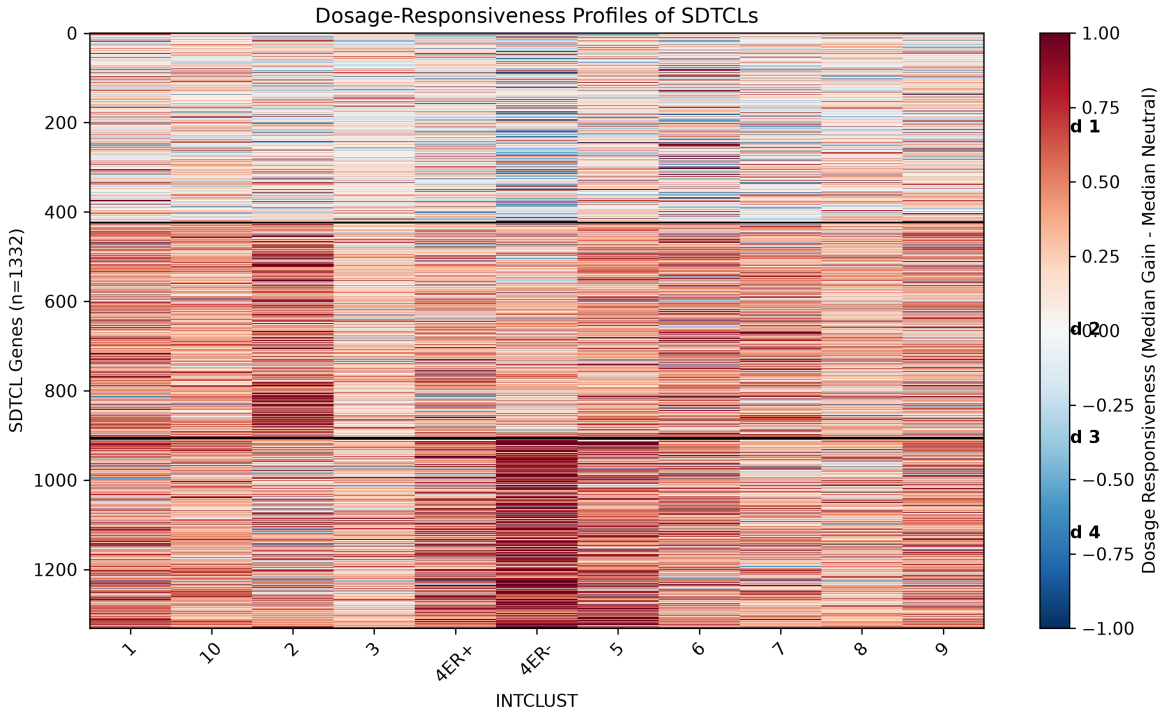


Figure 1: Heatmap illustrating the dosage-responsiveness profiles of Subtype-Divergent Coupling Loci (SDTCLs) across `INTCLUST` molecular subtypes. The color scale represents dosage responsiveness, calculated as the difference in median expression between copy-number gained and neutral samples. Hierarchical clustering partitions the genes into four regulatory modules, separated by horizontal black lines. The top section corresponds to Module 1 (the Repressive Module), comprising 424 genes characterized by significantly lower mean dosage-responsiveness (0.0875) compared to the strong dosage coupling observed in Modules 2 and 4 (mean responsiveness of 0.3823 and 0.4609, respectively). This visualizes the divergence in responsiveness across subtypes, demonstrating a localized "dosage buffering" effect where the expected transcriptional gain following copy-number amplification is systematically suppressed for specific genomic loci.

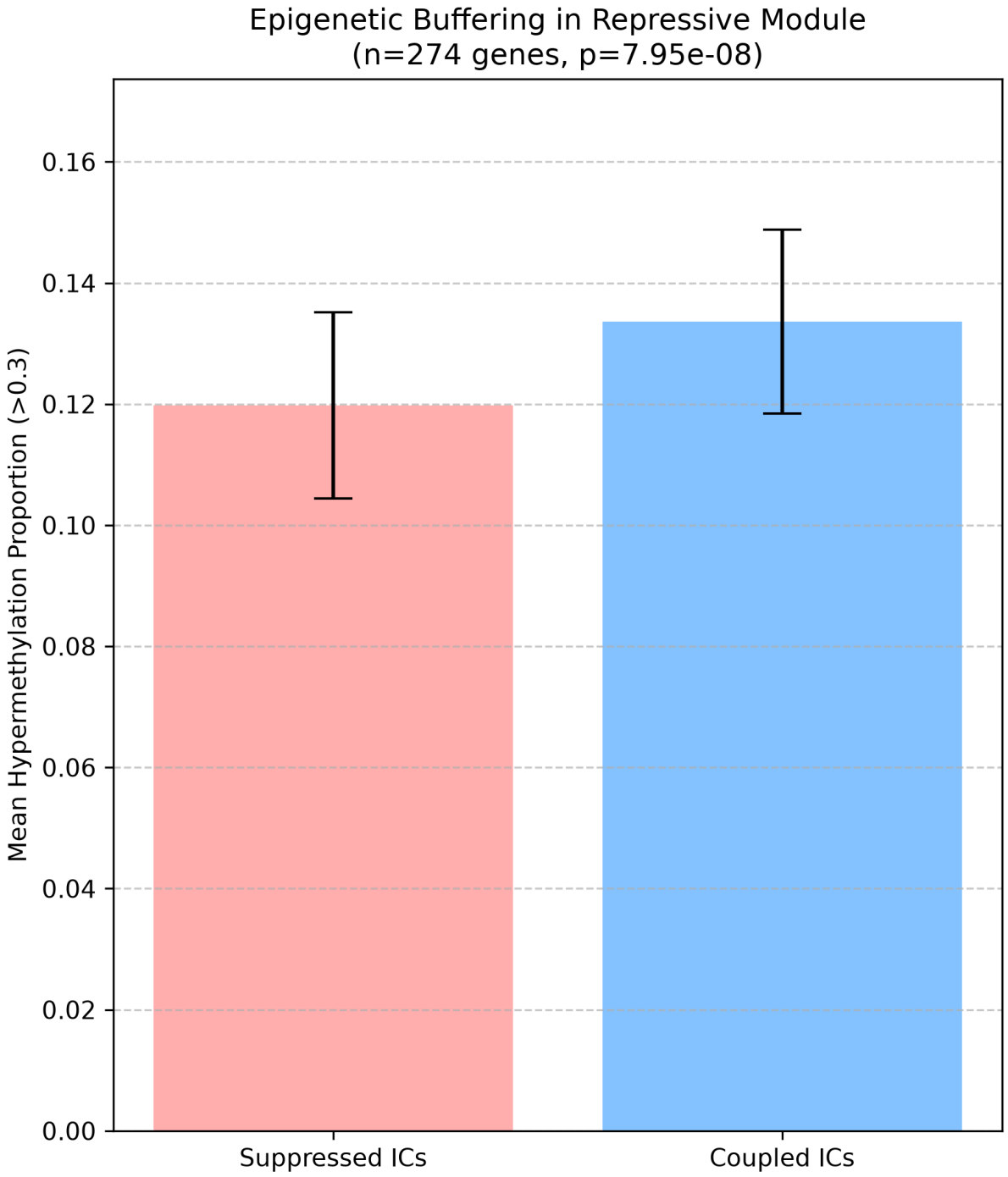


Figure 2: Comparison of mean promoter hypermethylation proportions ($\beta > 0.3$) for genes within the Repressive Module between suppressed and coupled `INTCLUST` groups. The bar plot demonstrates that the mean hypermethylation proportion is significantly lower in suppressed groups (0.1198) compared to coupled groups (0.1336; $p = 7.95 \times 10^{-8}$). Although statistically significant, this paradoxically directed association reveals that widespread promoter hypermethylation is not the primary driver of the dosage-buffering effect, suggesting that alternative mechanisms, such as chromatin state or post-transcriptional regulation, are likely responsible for the observed transcriptional dampening.

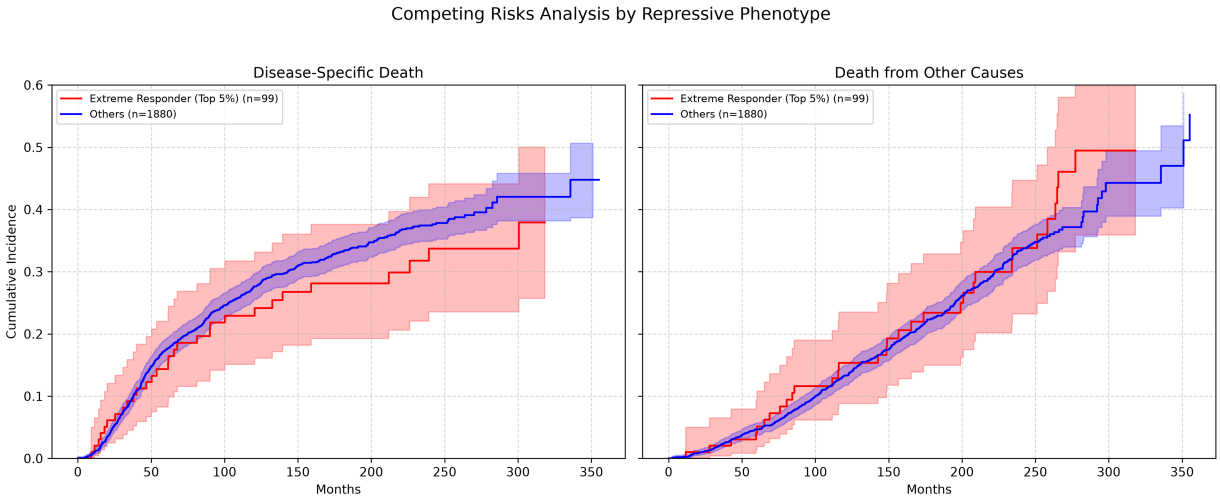


Figure 3: Non-parametric Aalen-Johansen cumulative incidence curves for disease-specific death and death from other causes, stratified by the Extreme Responder phenotype. Patients are grouped into Extreme Responders, representing the top 5% of the cohort with the highest repressive burden (red, $n = 99$), and the remaining patients (blue, $n = 1880$). The overlapping curves and shaded confidence intervals reveal no significant separation between the groups for either clinical endpoint. This demonstrates that the Extreme Responder phenotype, which represents maximal transcriptional buffering, provides no independent prognostic information for patient survival.

3.4 Cohort-wise replication and meta-analysis

To rigorously test the structural stability of the observed null effect, we performed independent Cox modeling within each of the five molecular accrual cohorts. The lack of association was highly consistent across the partitions. For instance, Cohort 3 showed no significant effect (HR = 0.775, $p = 0.309$). While Cohort 5 exhibited a nominally significant positive association (HR = 3.115, $p = 0.045$), this localized effect failed to replicate in any of the larger cohorts.

A random-effects inverse-variance weighted meta-analysis of the five cohorts yielded a pooled Hazard Ratio of 1.013 (95% CI: 0.606–1.693, $p = 0.962$) with moderate heterogeneity ($I^2 = 22.1\%$). The corresponding forest plot (Figure 4) visually summarizes these findings, confirming that the repressive phenotype lacks a reproducible impact on tumor lethality. Ultimately, despite the robust statistical identification of subtype-divergent dosage buffering at the molecular level, this pervasive transcriptional rewiring does not translate into an independent axis of clinical vulnerability.

4 Conclusions

In this study, we systematically mapped localized transcriptional rewiring across 1,980 breast cancer genomes to evaluate the clinical implications of copy-number dosage buffering. By modeling the interaction between categorical copy-number alterations and molecular subtypes, we identified 424 subtype-divergent coupling loci that constitute a distinct repressive module. Within this module, the expected transcriptional increases following genomic amplification are systematically dampened in a subtype-dependent manner, confirming that the relationship between gene copy number and transcript abundance is highly context-specific.

Our investigation into the regulatory mechanisms underlying this dosage buffering revealed that promoter hypermethylation is paradoxically lower in suppressed groups. This finding demonstrates that widespread DNA methylation is not the primary driver of the observed transcriptional dampening. Instead, it suggests that alternative regulatory mechanisms, such as broader chromatin state alterations or post-transcriptional control, are likely responsible for the localized buffering of these loci.

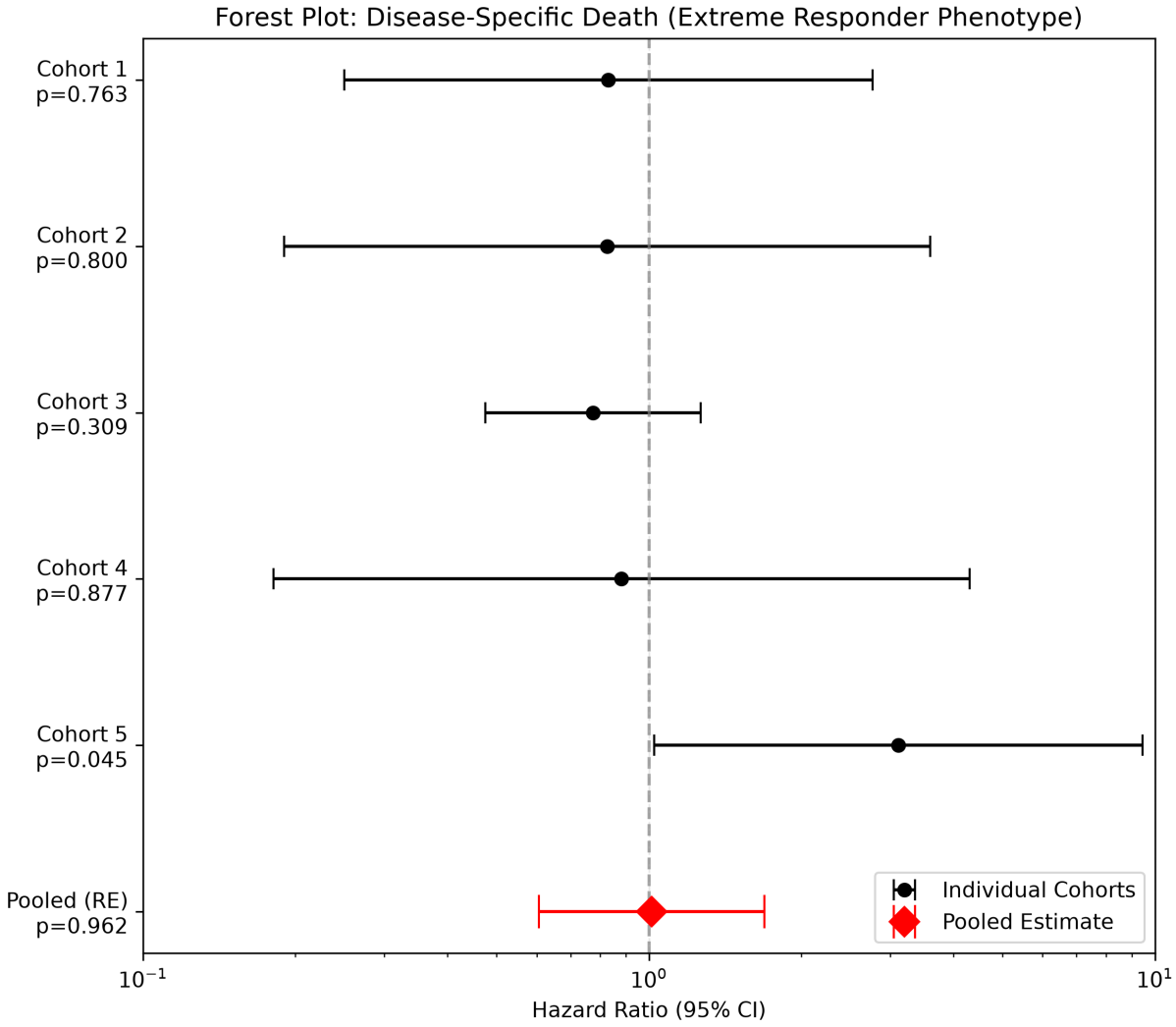


Figure 4: Forest plot summarizing the random-effects meta-analysis of disease-specific death associated with the Extreme Responder phenotype across five METABRIC accrual cohorts. Hazard ratios (black circles) and 95% confidence intervals are displayed for each cohort alongside the pooled estimate (red diamond). Although Cohort 5 shows a nominally significant positive association ($p = 0.045$), the pooled hazard ratio of 1.013 (95% CI: 0.606–1.693, $p = 0.962$) reveals no overall significant effect. This demonstrates that the repressive phenotype, representing maximal transcriptional buffering, lacks a reproducible impact on tumor lethality and provides no independent prognostic information for survival.

Crucially, we sought to determine whether this pervasive molecular rewiring translates into an independent axis of clinical vulnerability. By defining an extreme responder phenotype that captures the maximal burden of transcriptional buffering, we evaluated patient outcomes using cause-specific survival models. Despite the robust molecular presence of dosage buffering, our analyses revealed that the extreme responder phenotype provides no independent prognostic value for disease-specific death or death from other causes. This null finding proved structurally stable across five independent molecular accrual cohorts, confirming that the lack of clinical impact is a reproducible feature rather than a statistical artifact.

Ultimately, these results highlight a significant disconnect between localized molecular regulatory rewiring and clinical outcomes in breast cancer. While subtype-divergent dosage buffering is a statistically robust and biologically intriguing feature of genomic instability, it does not inherently constitute a reproducible axis of tumor lethality beyond established clinical and molecular prognostic markers.